Calibration of Trust and Distrust

by:

Barbara D. Adams, Ph.D.
Human*systems*® Incorporated
111 Farquhar St., 2nd floor
Guelph, ON N1H 3N4

Project Manager:
Barbara D. Adams, Ph.D.
(519) 836 5911

DRDC Scientific Authority
Carol McCann
(416) 635-2190

PWGSC Contract No. W7711-017747/001/TOR
Call-Up 7747-20

The scientific or technical validity of this Contract Report is entirely the responsibility of the contractor and the contents do not necessarily have the approval or endorsement of Defence R&D Canada.

Defence Research & Development Canada-Toronto

Contract Report

CR 2005-232

2005-11-02

106-02490/

Author

Barbara D. Adams, Ph.D.

Approved by

Carol McCann
Head, Command Effectiveness & Behaviour

Approved for release by

K.M. Sutton
Chair, Document Review and Library Committee

© HER MAJESTY THE QUEEN IN RIGHT OF CANADA (2005)
as represented by the Minister of National Defence

© SA MAJESTE LA REINE EN DROIT DU CANADA (2005)
Defense Nationale Canada

Abstract

Calibration is posited to play a key role in the development and maintenance of trust and distrust, but current literature gives little clear direction on exactly how this calibration might occur. This paper proposes a preliminary model of trust calibration and explores the process in detail. More specifically, trust calibration is initiated when mismatches occur between one's expectations about the outcome of a trust decision in comparison to the actual outcome. At this point, sensemaking or an attempt to understand the violation occurs, after which responsibility for the violation is assigned. When another person is seen as culpable, an assessment of intentionality is undertaken, in which the motives of the trustee are assessed. Seeing a trustee as both responsible and as having malevolent intention sets the stage for distrust, whereas viewing this person as responsible but not malevolent simply lessens existing trust. Other factors that impact on trust calibration processes are also considered.

Résumé

On postule que l'évaluation joue un rôle clé dans le développement et le maintien de la confiance et de la méfiance, mais la littérature actuelle donne peu d'indications claires sur la façon dont se fait cette évaluation. Cette étude propose un modèle préliminaire d'évaluation de la confiance et examine le processus en détail. L'évaluation de la confiance s'amorce plus précisément lorsque les attentes par rapport à une décision de faire acte de confiance ne concordent pas avec le résultat réel de cette décision. C'est alors qu'intervient la recherche de sens ou l'effort visant à comprendre le bris de confiance, suivi de l'attribution de la responsabilité de cette violation. Lorsqu'un coupable est désigné, une évaluation de l'intentionnalité est entreprise, laquelle comprend une évaluation des motifs du dépositaire de la confiance. Si cette personne est considérée à la fois comme responsable et comme habitée de mauvaises intentions, la méfiance s'installe, tandis que si elle est jugée responsable mais dénuée de mauvaises intentions, la confiance diminue tout simplement. D'autre facteurs influant sur l'évaluation de la confiance sont également étudiés.

Executive Summary

Although calibration is posited to play a key role in the development and maintenance of trust and distrust, current literature gives little clear direction on exactly how this calibration might occur. This paper proposes a preliminary model of trust calibration and explores the process in detail with respect to both a violation involving a previously trusted person (positive summary judgement) and a previously distrusted person (negative summary judgement). Trust calibration is initiated when mismatches occur between one's expectations about the outcome of a trust decision in comparison to the actual outcome of the trust decision. With little mismatch, the pre-existing view of the trustee will be relatively unaltered. However, when larger mismatches occur, individuals are likely to initiate sensemaking processes, in which they attempt to understand why the violation occurred. This sensemaking also includes the assignment of responsibility for the violation. When one accepts responsibility for the violation (rather than blaming the trustee), this is likely to initiate changes in one's own approach to trust issues. When the trustee is seen as culpable, an assessment of intentionality is undertaken, in which the motives of the trustee are assessed. Seeing a trustee as both responsible and as having malevolent intention sets the stage for distrust, whereas viewing this person as responsible but not malevolent simply lessens existing trust. Other factors that impact on trust calibration processes are also considered.

Sommaire

Même si l'on postule que l'évaluation joue un rôle clé dans le développement et le maintien de la confiance et de la méfiance, la littérature actuelle donne peu d'indications claires sur la façon dont se fait cette évaluation. Cette étude propose un modèle préliminaire d'évaluation de la confiance et examine le processus en détail par rapport à une violation mettant en cause une personne à qui l'on faisait auparavant confiance (jugement sommaire positif) et une personne dont on se méfiait auparavant (jugement sommaire négatif). L'évaluation de la confiance s'amorce lorsque les attentes par rapport à une décision de faire acte de confiance ne concordent pas avec le résultat réel de cette décision. Si le décalage est mince, la vision antérieure du dépositaire de la confiance demeurera sensiblement la même. Mais s'il est plus important, il est probable que s'enclenche un processus de recherche de sens visant à comprendre le pourquoi de la violation et à lui attribuer un responsable. Lorsque la personne qui fait confiance accepte la responsabilité de la violation (plutôt que de jeter le blâme sur le dépositaire de sa confiance), elle a de fortes chances de modifier son approche à l'égard de la confiance. Lorsqu'un coupable est désigné, une évaluation de l'intentionnalité est entreprise, laquelle comprend une évaluation des motifs du dépositaire de la confiance. Si cette personne est considérée à la fois comme responsable et comme habitée de mauvaises intentions, la méfiance s'installe, tandis que si elle est jugée responsable mais dénuée de mauvaises intentions, la confiance diminue tout simplement. D'autre facteurs influant sur l'évaluation de la confiance sont également étudiés.

Table of Contents

ABSTRACT I

RÉSUMÉ I

EXECUTIVE SUMMARY II

SOMMAIRE II

TABLE OF CONTENTS III

1. CALIBRATION OF TRUST AND DISTRUST 1

2. REFERENCES 7

HUMANSYSTEMS

1. Calibration of Trust and Distrust

The purpose of this short paper is to describe the process by which trust and distrust are calibrated. The issue of trust calibration plays a key role in our developing work on trust and distrust in small military teams (Adams and Webb, 2003; Adams and Sartori, 2005). We argue that people learn from their trust-related experiences and actively alter their judgements of the trustworthiness of others through calibration, typically in response to trust violations. The process of calibration is important in trust development in general, but is particularly important for clarifying how distrust develops over time. However, existing literature provides very little specific description of exactly how trust calibration might occur.

The purpose of this paper is to explore the concept of trust calibration in more detail. Unfortunately, although many trust theorists have discussed the issue of calibration (e.g. Mayer, Davis and Schoorman, 1995), there has been very little direct coverage of it in the literature. As such, although guided as much as possible by the available literature, these ideas are speculative and in need of validation.

The term "calibration" (Webster's Dictionary) has been defined as:

> "The act of checking or adjusting (by comparison with a standard) the accuracy of a measuring instrument; (e.g. "the thermometer needed calibration")."

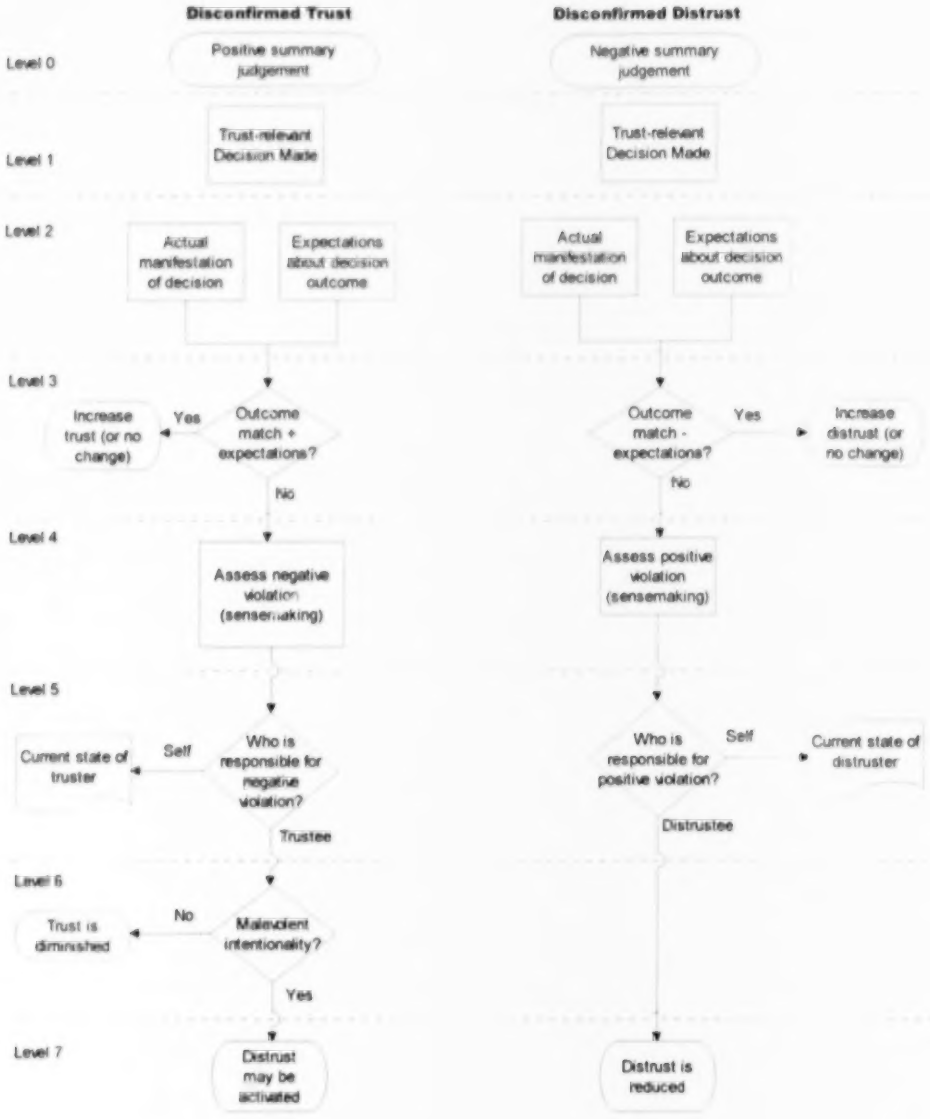
This definition is in keeping with informal usage of the term "calibration", which assumes some sort of adjustment to an existing standard. For example, this term is frequently employed in industrial applications, where the performance of one system can be compared to and then calibrated with the performance of another.

In the psychology literature, the term "calibration" is used in several different contexts. One usage of the term also involves the comparison to an objective standard. In the behavioural decision making literature, for example, Weber and Brewer (2003) suggest that

> "...calibration refers to the extent to which participants' judgements of confidence in their decisions (made on the percentile scale) correspond with the actual probability that their decision was correct".

In this work, the term "calibration" refers to the match between participants' confidence and actual judgement accuracy. A better match suggests that one's judgement processes are better calibrated to the optimal decision. In other behavioural decision making research, calibration is defined as the extent to which expected probability equals observed probability in the long run (Winkler and Murphy, 1968; cited in Lichtenstein, Fischhoff, and Phillips, 1982).

Our proposed model of trust and distrust calibration (see Figure 1), however, depicts the trust calibration process as somewhat more complex than simply a comparison to an existing standard. Although trust calibration processes can be either discrete or continuous, the following flow chart depicts it with respect to a discrete violation.



Assumes attribution of personal blame to the trustee or distrustee can be made
Assumes established unidimensional summary judgements

Figure 1. The Process of Trust Calibration

Figure 1 depicts the process of calibration after a trust violation which begins with a summary or baseline judgement (Level 0).[1] Summary judgements about a given person represent the logical culmination of many different judgements about trustworthiness in many different domains. Further, this summary judgement may not be wholly internally consistent and may contain both elements of trust as well as distrust. However, this global summary judgement is the best compromise of all the available information that speaks to either trustworthiness or distrustworthiness.

This flow chart also assumes (see Level 1) that some sort of decision to either trust or distrust needs to be made. As indicated in previous work (Adams and Webb, 2003), this decision is influenced by the need to trust, as well as by the perceived level of risk (e.g. cost and benefit analysis) within the situation.

Level 2 shows two factors that influence the early stage of calibration. Specifically, at this stage, there is an actual result of the trust decision.[2] One's decision to put faith in another teammate is either rewarded or invalidated. This outcome stands against one's expectations about the decision to place trust in another person. Of course, expectations about the trustworthiness or distrustworthiness of another teammate stem from the overall summary judgement. Once the outcome of the decision is known, the trustor (or distruster) will compare the expected outcome to the actual outcome, and will make a decision about how well they match.

Level 3 shows the comparison between the expected outcome and the actual outcome. When there is a match between actual and expected outcomes, calibration may not be necessary. For example, when positive expectations have been confirmed, trust may increase or be relatively unaffected. Similarly, when one has a pre-existing negative summary judgement that is confirmed, this may promote some increase in distrust, or may simply be incorporated into the summary view with little change in one's negative expectations. As Level 3 suggests, a calibration process may be initiated when one notes a mismatch between one's current view of another person and one's expectations. Unlike other forms of calibration within the interpersonal domain, however, there is often no clear external or apriori standard by which to judge the actual or absolute trustworthiness of another person. Sometimes, people can make absolute comparisons with respect to an external or *a priori* standard (e.g. of how trustworthy people normally behave). Social norms, for example, provide information about trustworthy and untrustworthy behaviour and may provide a standard by which to judge the actions of others. When clear and absolute trustworthiness standards do not exist, relative standards must be used in order to gauge the trustworthiness of those around us. In these cases, a comparison can be made by matching one's prior expectations (and at a broader level, one's view of this person) with the outcome of one's trust decision.

In Level 4, when positive apriori expectations have been violated, the victim now undertakes the process of assessing or making sense of the violation. Substantive mismatches may activate a concerted process that attempts to reconcile the new information into the summary judgment. The extent of sensemaking efforts will depend on the discrepancy between the expected and the actual outcome, with highly discrepant effects being more likely to invoke change. For example, a negative or malevolent violation committed by a person previously seen as trustworthy is simply more surprising, and hence more likely to set sensemaking processes into play than is positive and supportive behaviour on the part of a previously distrusted person. This difference is indicated by the relative sizes of the sensemaking boxes. The extent of calibration that must occur within a given

---

[1] In order to depict both positive and negative calibration processes, the left side of the flow chart shows disconfirmed trust, and the right side shows disconfirmed distrust.

[2] Of course, the outcome of a trust decision is not necessarily discrete and transparent, and may or may not be immediately known. When it is not known, trust calibration may not occur.

situation will also depend on many of the same person-based, situational and contextual factors that influence both trust (Adams, Bryant and Webb, 2001) and distrust (Adams and Sartori, 2005). For example, qualities of the truster or trustee (such as one's goals at the time, the importance of these goals, etc.) will influence the extent to which discrepancies are attended to, and are likely to evoke change in one's overall summary judgement. External pressure to have properly calibrated trust-relevant expectations (e.g. when a leader must complete a performance appraisal) will also make one more likely to attend to discrepancies and to adjust when discrepancies between one's expectations and actual outcomes occur.

As indicated in Level 5, the need to assess the violation further typically culminates in attempting to ascribe responsibility to the appropriate party (Adams and Sartori, 2005). A key issue here is where responsibility for the violation is assigned. If one assigns responsibility to oneself, one may initiate changes in how the trustworthiness of others is judged. This may, for instance, result vowing to never trust anyone again, or to look for factors within prospective "trustees" that may have influenced one's error. We might remember that we had a bad feeling but did not heed it. Whatever the issue, a person can alter him or herself, and no additional calibration occurs with respect to this specific incident or with specific reference to the target involved in the violation. When responsibility is assigned to the trustee, however, this may entail changes in one's perceptions of specific qualities or characteristics (e.g. a person is seen to be less competent), or in the summary judgement of perceived trustworthiness.

As indicated in Level 6, when responsibility is assigned to the trustee, a decision about the malevolence of the act may be required.[3] When responsibility for the mismatch between one's expectations and the actual outcome is assigned, this attribution process can result in either a contextualized judgement being made (e.g. this person harmed me, but they had no choice, other contextual factors impacted etc.) or in the person being held responsible for the violation. Attributing malevolent intention to the violator makes it more likely that responsibility will be assigned at the dispositional rather than the situational level. Obviously, dispositional attributions strike closer to the core of who the violator is. This argument is consistent with research described in our distrust review, in which revenge fantasies depended on the attribution of responsibility, whether the act was seen as personal or not, whether one attributed the act to a role violation, or to the system within which the violation occurred (Bies and Tripp, 1996).

As indicated in Level 7, then, when a previously trusted person can be held responsible for a violation of trust with malevolent intentionality, the prototypical form of distrust may result. On the other hand, when a previously distrusted person acts in an unexpectedly positive way, this is likely to simply reduce distrust, but distrust would not typically be overturned. This parallels a common notion in the literature that trust is easier to build than to rebuild (e.g. Lewicki and Bunker, 1995).

Several related factors will also influence trust calibration processes. For example, the effect of a trust violation will be influenced by the complexity of the knowledge and belief structure associated with the violator. In keeping with the notion of link multiplexity advanced by Lewicki, McAllister and Bies (1998), if one's view of another teammate is determined by a multiplicity of different aspects (e.g. integrated knowledge over time), there will be more resistance to change, and the number of layers protecting the core summary judgement will be higher. As such, for a violation to reach closer to the core (even just beginning to impact on one's attribution of intentionality), violations committed by this person would have to be either very frequent, very influential, or to provide irrefutable evidence of malevolent intent. This might occur, for example, with a teammate having been caught in a blatant lie which they were shown to have knowingly perpetuated. When the core summary judgement about another person as being essentially trustworthy is challenged, this may signal the movement from the realm of trust to the realm of distrust.

---

[3] The issue of malevolence is discussed in detail in Adams and Sartori (2005).

Another way to understand this process relates to coherence within belief systems (Thagard, 1992). As we argued in our first trust review (Adams et al., 2001), trust expectations help to form a small part of one's overall belief system with respect to another person. Our views of other people are likely to consist of many different facets, some positive, some negative, but arranged around a core belief about who this person really is. Although we may have many diverse expectations about another person, summary judgements about trustworthiness bring coherence to an otherwise potentially muddled picture and help us to understand how other people are likely to behave in many diverse contexts. According to our model of trust in teams, trust judgements are influenced by many different components, including a trustee's perceived competence, benevolence and integrity (Adams and Webb, 2003). These beliefs are not isolated, but actively work to both influence and to constrain each other. As such, someone who is benevolent is also likely to have integrity, and less likely to have malevolent intention.

Most importantly, these beliefs are also likely to have differential effects on trust calibration processes. From a coherence perspective, violations of trust are likely to influence the summary judgement that we have of another person, but the basis for the violation (e.g. competence vs. integrity) will also influence the speed at which this erosion occurs. By their very nature, violations of integrity are typically likely to be more influential on maintaining a coherently positive view of another person because integrity is perhaps more freely chosen (and hence, subject to the ascription of intentionality). For example, seeing a fellow teammate deliberately represent another person's idea as his own has the potential to create chaos within one's previously positive and coherent view of this person.[4] This kind of a violation shows deliberate malice and intent to harm another person. Even a single malevolent act against a fellow teammate that is interpreted to be intentional has the potential to alter one's view from having a trusted teammate to having a distrusted teammate. On the other hand, simple task competence may not have a great deal of impact on other trust-relevant characteristics, as it may require very little sensemaking activity to understand and to be able to dismiss. In such cases, a simple adjustment on the same active continuum that drives the summary judgement may be adequate (i.e. one's overall view is still that this person is trusted) and only a minor revision or downgrade of trust is necessary. In general, violations of integrity and benevolence are likely to have more impact on summary judgements than will competence.

This suggests that violations that impose more challenges on the coherence of previous belief systems will be more influential in initiating trust calibration processes. Summary judgements will only be used if the various elements that comprise the summary judgement actually cohere, and are able to explain the events that they encounter in the course of actually interacting with this person. When one's beliefs about who another person is are threatened by piecemeal evidence, continuous calibration efforts may need to be undertaken in order to sustain one's dominant expectations. However, when these beliefs can no longer be sustained, whole scale calibration may need to occur. As noted in our distrust review (Adams and Sartori, 2005), trust calibration is of critical importance in thinking about trust and distrust within military teams, and our review argues that trust and distrust (and their correlates) have both positive and negative elements. The key, it seems, is properly calibrated trust or distrust that changes as the circumstances require. For small military teams, the goal is not necessarily to promote trust and to wholly discourage distrust, but to promote flexibility in responding to change while specifically discouraging highly entrenched beliefs and behaviours, regardless of whether they are positive or negative (Deutsch, 1973). Any proclivity toward either an exaggerated or rigid use of trust/distrust has the potential to be problematic within the military context.

---

[4] It is important that coherence not be equated with consistent positivity or negativity. It is possible to have both within a summary judgement, but they work together to constrain and explain each other. It is also important to note that coherence is not intended as a product of purely cognitive processes. Clearly, affect and emotion will also influence coherence (Thagard, 1992).

This suggests that within military teams, proper trust calibration is a very important issue that needs to be understood in more detail. The preliminary trust calibration model proposed in this paper represents a start toward this goal.

2. References

ADAMS, B.D., SARTORI, J. (2005). Dimensionality of Trust Literature Review. *Draft report in preparation for Department of National Defence.*

ADAMS, B.D., WEBB, R.D.G. (2003). Model of Trust Development in Small Teams. *Report to Department of National Defence. DRDC No. CR-2003-016.*

ADAMS, B.D., BRYANT, D.J., WEBB, R.D.G. (2001). Trust in Teams Literature Review. Report to Department of National Defence. *Report No. CR-2001-042.*

BIES, R., & TRIPP, T. M. (1996). Beyond Distrust: "Getting Even" and the Need for Revenge. In R. Kramer & T. Tyler (Eds.), Trust in Organizations: Frontiers of Theory and Research (pp. 246-260). Thousand Oaks, CA: Sage Publications.

DEUTSCH, M. (1973). The resolution of conflict: Constructive and destructive processes. New Haven, CN: Yale University Press.

KIM, P. H., FERRIN, D.L., COOPER, C., & DIRKS, K. (2004). Removing the Shadow of Suspicion: The effects of apology versus denial for repairing competence- versus integrity-based trust violations. Journal of Applied Psychology, 89(1), 104-118.

LEWICKI, R., & BUNKER, B. (1995). Trust in relationships: A model of development and decline. In Bunker, B., & Rubin, J. (Eds.). *Conflict, cooperation, and justice: Essays inspired by the work of Morton Deutsch.* (pp.133-173). San Francisco, CA, US: Jossey-Bass Inc.

LEWICKI, R., MCALLISTER, D and BIES, R. (1998). Trust and Distrust: New Relationships and Realities. Academy of Management Review, 23(3): 438-445.

LICHTENSTEIN, S., FISCHHOFF, B., PHILLIPS, L. (1982). Calibration of Probabilities: The state of the art to 1980. In Judgment Under Uncertainty: Heuristics and Biases (D. Kahneman, P. Slovic, A. Tversky, eds.), pp. 306-34. Cambridge, England: Cambridge University Press.

MAYER, R., DAVIS, J., & SCHOORMAN, F. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709-734.

THAGARD, P. (1992). *Conceptual Revolutions*. Princeton: Princeton University Press.

WEBER, N. and BREWER, N. (2003). The effect of judgment type and confidence scale on confidence-accuracy calibration in face recognition. Journal of Applied Psychology, 88(3), 490-499.

DOCUMENT CONTROL DATA
(Security classification of the title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.)

Publishing: DRDC Toronto
Performing: Humansystems® Incorporated
Monitoring:
Contracting: DRDC Toronto

2. SECURITY CLASSIFICATION
(Overall security classification of the document including special warning terms if applicable.)

UNCLASSIFIED

3. TITLE (The complete document title as indicated on the title page. Its classification is indicated by the appropriate abbreviation (S, C, R, or U) in parentheses at the end of the title)

Calibration of Trust and Distrust (U)

4. AUTHORS (First name, middle initial and last name. If military, show rank, e.g. Maj. John E. Doe.)

Barbara D. Adams; Carol McCann

5. DATE OF PUBLICATION (Month and year of publication of document.)	6a NO. OF PAGES (Total containing information, including Annexes, Appendices, etc.)	6b. NO. OF REFS (Total cited in document.)
September 2005	15	12

7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Contract Report

8. SPONSORING ACTIVITY (The names of the department project office or laboratory sponsoring the research and development – include address.)

Sponsoring:
Tasking:

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant under which the document was written. Please specify whether project or grant.)	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)
10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document) DRDC Toronto CR 2005-232	10b. OTHER DOCUMENT NO(s). (Any other numbers under which may be assigned this document either by the originator or by the sponsor.)

11. DOCUMENT AVAILABILIY (Any limitations on the dissemination of the document, other than those imposed by security classification.)

Unlimited distribution

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, when further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected.))

Unlimited announcement

DOCUMENT CONTROL DATA

(Security classification of the title, body of abstract and indexing annotation must be entered when the overall document is classified)

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

(U) Calibration is posited to play a key role in the development and maintenance of trust and distrust, but current literature gives little clear direction on exactly how this calibration might occur. This paper proposes a preliminary model of trust calibration and explores the process in detail. More specifically, trust calibration is initiated when mismatches occur between one's expectations about the outcome of a trust decision in comparison to the actual outcome. At this point, sensemaking or an attempt to understand the violation occurs, after which responsibility for the violation is assigned. When another person is seen as culpable, an assessment of intentionality is undertaken, in which the motives of the trustee are assessed. Seeing a trustee as both responsible and as having malevolent intention sets the stage for distrust, whereas viewing this person as responsible but not malevolent simply lessens existing trust. Other factors that impact on trust calibration processes are also considered.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

(U) Calibration; Trust; Distrust